MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 1: 1–13
DOI 10.3897/mbmg.2.26939

Research Article

A meta-barcoding census of freshwater planktonic protists in Appalachia – Natural Tunnel State Park, Virginia, USA

A. Bruce Cahoon[1], **Ashley G. Huffman**[1], **Megan M. Krager**[2,3], **Roseanna M. Crowell**[1]

1 *Department of Natural Sciences, University of Virginia's College at Wise, 1 College Ave., Wise, Virginia 24293, USA.*

2 *Natural Tunnel State Park, 1420 Natural Tunnel Pkwy., Duffield, Virginia 24244, USA.*

3 *Bays Mountain Park, 853 Bays Mountain Rd, Kingsport, TN 37660, USA.*

Corresponding author: *A. Bruce Cahoon* (abc6c@uvawise.edu)

Academic editor: *J. Pawlowski* | **Received** 24 May 2018 | **Accepted** 11 September 2018 | **Published** 10 October 2018

Abstract

The purpose of our study was to survey the freshwater planktonic protists within an inland natural preserve in the Ridge and Valley physiographic province of the Appalachian Region using metabarcoding. Microbial eukaryotes are essential primary producers and predators in small freshwater ecosystems, yet they are often overlooked due to the difficulty of identification. This has been remedied, in part, by the cost reduction of high throughput DNA sequencing and the growth of barcode databases, making the identification and analysis of microorganisms by way of metabarcoding surveys in complex ecosystems increasingly feasible. Water samples were collected from five sites at the Natural Tunnel State Park in Scott County, VA (USA), representing three common bodies of water found in this region. Samples were initially collected during a Bioblitz event in April 2016 and then seven and fourteen weeks afterwards. Metabarcode analysis of the 23S and 18S genes identified 3663 OTUs representing 213 family level and 332 genus level taxa. This study provides an initial barcode census within a region that has a reputation as a temperate biodiversity "hotspot". The overall protist diversity was comparably high to other temperate systems, but not unusually high; the microalgal diversity, however, was higher than that reported for other temperate regions. The three types of water bodies had their own distinctive protist biomes despite close proximity.

Key Words

Metabarcoding, microbial eukaryotes, algae, Bioblitz

Introduction

The region of the Cumberland Mountains and Ridge and Valley physiographic provinces of Appalachia, located at the intersection of Virginia, Kentucky, Tennessee and North Carolina, is considered one of the most biodiverse temperate regions in North America and is home to a large number of endemic freshwater macro-invertebrates (Neves et al. 1997; Parmalee and Bogan 1998). Despite this recognition, there has been very little focus on the microbiome of these water systems. The purpose of this study was to address the question; 'How diverse is the eukaryotic microbiome of the small freshwater systems in the limestone Ridge and Valley region of Appalachia?'

Microbial eukaryotes (protists) are the most abundant eukaryotes on Earth (Zhao et al. 2011) and are essential components of aquatic systems. Constituting much of the primary production within those food webs, protists have a large collective photosynthetic carbon fixation potential (Moss 2017) and are important in nutrient cycling as they graze on microbes thus stimulating the turnover of essential nutrients that may be sequestered in prokaryotic and/or eukaryotic biomass (Worden et al. 2015, Falkowski et al. 2008, Finlay and Esteban 1998). Protists also serve as essential prey for small metazoans (Jack and Gilbert 1997, Stoecker and Capuzzo 1990).

Morphological surveys have uncovered an astounding inventory of freshwater protists (John 2002,

Copyright *A. Bruce Cahoon et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

Wehr et al. 2015) and, prior to the application of molecular survey methods, some hypothesised that most species had been discovered (Finlay 1998). This was overturned with the application of molecular survey approaches using Sanger-based sequencing (López-García et al. 2001, Moon-van der Staay et al. 2001, Moreira and López-García 2002) that could identify rare species and those recalcitrant to culturing. When applied to freshwater protists, diversity was found to be far from completely described (Šlapeta et al. 2005). This approach has now evolved to use high throughput metabarcode sequencing that provides a set of tools to identify an even greater assemblage of species, including ones that are in extremely low abundance, aka the rare microbiome (Medinger et al. 2010). These advantages, coupled with rapidly expanding barcode databases, are providing new opportunities to identify and quantify microbes in various understudied ecosystems. The majority of published metabarcoding surveys have focused on the prokaryotic microbiome (Grattepanche et al. 2014) or marine protists (example: Stoeck et al. 2010) but there are a growing number of published surveys of temperate freshwater systems that have shown them to be extensive reservoirs of protist biodiversity (Lefranc et al. 2005, Zalack et al. 2006, Nolte et al. 2010, Debroas et al. 2015, 2017, Lepère et al. 2013, Mangot et al. 2012, Stoeck et al. 2014, Bricheux et al. 2013, Brown et al. 2015, Simon et al. 2015a, 2015b). These studies have provided evidence of complex microeukaryotic biomes with rare taxa (Aguilar et al. 2016; Debroas et al. 2015). In addition, they have found organisms thought to reside exclusively in marine ecosystems (Simon et al. 2015a), others never found in other surveys (Lepère et al. 2013) and have shown that geographically close water systems can have very different microbiomes (Simon et al. 2015b).

This paper is, to our knowledge, one of the few surveys of microbial eukaryotes undertaken in the Appalachian region of North America and the first application of eukaryotic microbiome metabarcoding in a Bioblitz event. A Bioblitz uses a combination of experts and citizen scientist volunteers to survey a natural area for pre-defined groups of organisms over the course of 24 hours. The survey presented in this paper provides an inventory of 3663 OTUs representing 213 family level and 332 genus level taxa from five sample sites within a state park. These data were used to test two hypotheses: (i) overall diversity would be high; perhaps higher than comparable temperate regions and (ii) the five sample sites would have their own distinct protist profiles.

Materials and methods

Sample sites

Water samples were collected at five sites in the Natural Tunnel State Park (Scott County, VA, USA). These sites were chosen because they represent the most abundant types of natural water bodies found in this physiographic region: 1 – a lentic ephemeral pool in an abandoned Quarry. The floor of the quarry is relatively flat and occupies an area 13,500 m^2 (estimate calculated using Google Earth Pro, https://www.google.com/earth/desktop/). The pool was estimated to be ~400–500 m^2 and depth was ~0.5 m at its deepest point. 2 – Stock Creek is a stony stream running through agricultural areas and lined with riparian vegetation prior to entering the park where it is surrounded by deciduous forest. It is responsible for carving the tunnel from which the park derives its name. Stock Creek North (SC-N) samples were taken from a lotic area with no canopy cover occurring ~750 m upstream of the North Portal of the tunnel. At this site, the creek is 9–10 m wide and 1–1.5 m in depth. The broad flat bank on the eastern side of this sample site is a popular destination for anglers. 3 – Stock Creek South (SC-S) is about ~400 m downstream of the south portal of the tunnel. This site is shallow (≤1 m in depth) and 9–10 m wide with many riffle pools and extensive deciduous canopy cover. 4 –a stream feeding into Stock Creek with a small Waterfall, ~3 m in height, with extensive deciduous canopy cover, ~1–2 m wide, very shallow (5–10 cm deep) and running down a rocky slope where it enters Stock Creek. Samples were collected from the waterfall. 5 – a different stream feeding into Stock Creek near an abandoned settlement with a derelict Waterwheel with extensive deciduous canopy cover. This stream was 1–2 m wide and very shallow (5–10 cm depth). The Waterfall and Waterwheel streams are intermittent and fed by a combination of artesian sources and rainfall. They were free-flowing during our sampling period but they can dry up during periods of low rainfall. Photographs of each site are located in Suppl. material 1: Figure 1 while GPS coordinates and altitudes are listed in Suppl. material 3: Table 1. Canopy cover and water flow rates were not quantified. Samples were collected from shallow sites (streams and quarry) by hand and from the creek sites with a Van Dorn sampler. Three one-litre samples were collected at each site and time interval.

This sampling and identification project was included in the Natural Tunnel Park Bioblitz held on 22 April 2016. During the event, the authors collected water samples from pre-selected sites within the park and then engaged citizen scientist volunteers in morphological identifications using microscopy. The three water samples collected from each site were combined and cells were harvested by filtration on to 0.45 micron membranes using a disposable microfunnel with filter (Daigger & Co., Vernon Hills, IL, USA) and an electric vacuum pump (Welch, Mt. Prospect, IL, USA, Model 2534B-01A). This pore size was chosen since it should capture all but the smallest prokaryotes and picoplankton. All samples collected on 22 April were filtered on site and the filters maintained on ice for 4–6 hours before eventual storage at -20 °C. It has been observed that the protist community is highly dynamic with a large seasonal variation (Nolte et al. 2010; Stoeck et al. 2014). To account for this and to expand the total diversity seen in this study, replicate

samples were taken on 6 June and 20 July, which were approximately 7 and 14 weeks after the 22 April bioblitz. These samples were filtered in a laboratory within 2–4 hours of collection and filters stored at -20 °C.

DNA extraction, PCR, sequencing

DNA was extracted from cells collected on the filters using Qiagen's Power Water kit (Germantown, MD, USA). DNA was quantified using a Nanodrop Lite apparatus (ThermoFisher Scientific, Waltham, MA, USA).

The intention of this study was to sample as broadly as possible so primers were chosen that would theoretically identify the highest number of taxa. The ~500bp V4 region of the nuclear 18S rRNA, considered a universal protist barcode (Stoeck et al. 2010), was chosen to generally target protists due to its broad acceptance and usage in multiple barcoding surveys (Debroas et al. 2017, Pawlowski et al. 2012). The plastid 23S rRNA barcode was chosen for photoautotrophs (Sherwood and Presting 2007). It produces a ~400bp amplicon that has broadly identified photoautotrophic species in previous surveys (Stern et al. 2010; Pawlowski et al. 2012; Yoon et al. 2016; Cannon et al. 2016). Amplicons produced from a third primer set intended to target the diatom 18S V4 region (Zimmermann et al. 2011) were included in the sequencing lane but the list of OTUs and higher-level taxa obtained using these primers were nearly identical to those collected using the universal V4 18S primers, so these data were not used in this study.

Amplicons were produced from the environmentally derived DNA preparations using the primers listed in Suppl. material 4: Table 2, ExTaq polymerase (Takara, Shiga, Japan), Bio-Rad's C1000 thermal cycler (Hercules, California, USA) and thermal cycler programmes listed in Suppl. material 4: Table 2. Amplicons were produced, in duplicate, from each site, visualised by gel electrophoresis, removed from the gels using an x-traca gel extraction tool (Promega, Madison, WI, USA) and purified using the Zymoclean Gel DNA Recovery kit (Zymo Research, Irvine, CA, USA). The duplicate purified amplicons from each site/date were combined and used as a template for the addition of sample specific barcodes as well as adaptor sequences necessary for the Illumina sequencing process and multiplexing (protocol is listed in Suppl. material 4: Table 2). Amplicons were paired-end sequenced by a commercial sequencing facility (Genewiz, South Plainfield, NJ, USA) using an Illumina MiSeq system in a single unshared lane. All sequence data referenced in this paper are available through the National Center for Biotechnology Information (USA) Sequence Read Archive, accession PRJNA434596.

Bioinformatics pipeline, sorting of barcode sequences and OTU identification

Sequence information was delivered from Genewiz as de-multiplexed paired-end files from each sample site. Sequences were paired, sorted and Operational Taxonomic Units (OTUs) identified using a protocol developed for metabarcoding analysis using the commercial software Geneious (BioMatters Ltd, Auckland, New Zealand, http://www.geneious.com/tutorials/metagenomic-analysis, accessed December 2016). A total of 4,500,654 paired reads were produced using the three primer sets (Table 1 – Total Number of Paired Reads). Of these, 539,532 had Phred scores ≥30 and could be merged to create individual consensus sequences. Next, these sequences were compared to the primary primer sequences and only reads matching the primers with 100% similarity were retained, leaving 84,330 18S and 8,134 23S sequences (Table 1 – Sequences with a 100% primers match). These sequences were screened using GenBank's BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi accessed February–March 2017) to estimate how many sequences were potentially identifiable (Table 1 – Sequences with a GenBank Match). Sequences were aligned using a *de novo* assembly approach to remove redundancy (the reads were binned by similarity) and created a list of consensus sequences that could serve as the initial list of OTUs (Table 1 – Estimate of Unique OTU's via *de novo* Assembly). No more than 2% mismatches or gaps were allowed in the *de novo* alignment. This level of stringency meant the OTUs list still had some redundancy, but it was unlikely that unique sequences were combined into a single consensus. GenBank was searched again using the low-redundancy OTUs sequence list using the BLAST tool. Matching GenBank records were compiled and duplicates were removed to create a searchable database. The sequences from each site (Table 1 – Sequences with a 100% Primers Match) were screened against this database using the following parameters: a 95% or higher similarity was considered the same genus as the database record and a 90% or higher, the same family. These criteria were derived based on estimates of average species (98%) and genus (89%) sequence similarities reported by Caron et al. (2009) for the complete 18S rDNA and adjusted to increase stringency in an attempt to accommodate the V4 variable region barcode. Taxonomic evaluations were limited to the levels of genus and family since no single barcode can reliably resolve all groups at the species level (Pawlowski et al. 2012). All taxa that occurred as a single sequence read from all sites and dates were removed from the list since these could represent spurious reads. Taxonomic designations were initially based on GenBank records and then revised according to AlgaeBase (Guiry and Guiry 2017; http://www.algaebase.org, accessed June–July 2017).

Data analyses

To estimate if the number of sequences identified at each site provided the highest number of taxa possible, rarefaction was calculated and a graph generated using R software (R Core Team 2017) and the vegan package (Oskanen et al. 2017). In addition, C.hat (estimator of sample

Table 1. A summary of sequence reads, OTUs and taxon totals for each collection site and date and each barcode used in this study.

	Sequences	18S					23S				
	Total Number of Paired Reads	Sequences with a 100% Primers Match	Sequences with a GenBank Match	Estimate of Unique OTUs via *de novo* Assembly*	Identifiable Families	Identifiable Genera	Sequences with a 100% Primers Match	Sequences with a GenBank Match	Estimate of Unique OTUs via *de novo* Assembly*	Identifiable Algal Familes	Identifiable Algal Genera
Quarry – Week 0	380,808	10361	9923	2018	68	110	1317	995	309	16	18
Quarry -Week 7	148,452	6231	6144	1030	42	55	433	427	433	13	13
Quarry – Week 14	103,166	1160	1139	573	49	65	516	513	138	15	25
Stock Creek-North – Week 0	272,112	9293	8796	2313	60	95	821	725	130	16	18
Stock Creek-North – Week 7	477,600	5529	5264	2358	86	136	292	227	367	28	30
Stock Creek-North – Week 14	176,668	3370	3319	1200	52	75	229	216	72	12	12
Stock Creek-South – Week 0	238,990	4625	4257	1579	64	96	313	249	151	16	20
Stock Creek-South – Week 7	356,404	6946	6622	2675	83	129	335	294	146	18	23
Stock Creek-South – Week 14	312,426	7548	7396	2595	64	91	357	345	162	24	27
Waterfall –Week 0	298,584	5689	5396	1299	53	71	339	267	111	11	11
Waterfall –Week 7	254,362	4857	4607	1733	49	72	705	651	363	26	24
Waterfall –Week 14	382,956	5038	4769	1839	50	70	494	464	182	22	26
Waterwheel – Week 0	496,364	5225	4864	1667	65	102	1496	826	714	21	25
Waterwheel – Week 7	127,154	3176	3008	1363	47	63	125	121	70	10	9
Waterwheel –Week 14	474,608	5282	4970	2323	60	93	362	310	150	19	22
Column total [Σ]	4,500,654	84,330	80,474	26,565	892	1,323	8,134	6,630	3,498	267	303
Column average [μ] ± SD	300043.6 ± 128373.8	5622 ± 2313.6	5364.9 ± 2220.6	1771± 612.8	59.5± 12.7	88.2 ± 24	542.3 ± 393.5	442± 253.5	233.2 ± 174.8	17.8 ± 5.5	20.2 ±6.5
non-redundant total				3552	212	198			343	48	75

*The parameters used to match sequences for the initial OTU estimate were highly stringent. Due to this, some redundant sequences were categorized as separate OTUs which inflated the total number. Redundancies were identified and removed during library preparation and taxa screening.

coverage) scores were generated using iNEXT (Hsieh et al. 2016). Read counts were normalised by converting them to the relative proportion of reads from each collection site and time.

The Simpson's diversity index (*1-D*) was calculated according to Pielou (1969) and Evenness ($E_{1\text{-}D}$) calculated according to Smith and Wilson (1996). Shannon-Wiener diversity (*H'*) Shannon-Wiener evenness (E_{Heip}) and E_{var} were calculated according to Smith and Wilson (1996).

Multi-dimensional analyses were performed as Principle Coordinates analyses (PCO) with scatter plots to compare the taxonomic compositions of each sample site. These were generated using genera and sequence reads with the software package PRIMER-E with the PERMANOVA add-on (Clarke and Gorley 2015). For PCO, genus counts were normalised with the square-root function and resemblance estimated using the Bray-Curtis index.

Results

The eukaryotic microbiome (alpha-diversity)

OTU richness, as defined by sequence diversity for each sample, ranged from 573–2675 (μ=1771±612.8) for 18S and 70–714 (μ=233±174.8) for 23S (Table 1) with a non-redundant total of 3663. A total of 95.3% of the18S sequences and 81.5% of 23S were taxonomically identifiable. The 18S barcode identified 212 protist families and 198 genera while the 23S identified 48 families and 75 genera (Table 1). An overview of the number of identifiable OTUs (genus level) and sequence abundance for each barcode at each location and date are shown in Figure 1. Both barcodes used in this survey, plastid 23S rDNA and nuclear 18S rDNA, could identify photosynthetic autotrophic protists, aka algae, (Chlorophyta, Glaucophyta, Rhodophyta, Charophyta, Cryptophyta, Haptophyta, Ochrophyta, Bacillariophyta and some Dinoflagellata). The 18S barcode identified a much larger number of photoautotrophic genera OTUs (198) than 23S (75) but the genera from the two barcodes lacked complete concordance. Collectively, the two barcodes identified 220 genera: 22 genera (9.5%) were uniquely identified by 23S and 145 (65.9%) by 18S. Fifty-three genera (24.5%) were identified by both barcodes. A unified non-redundant list was created using results from both barcodes and used for all downstream analyses (Suppl. material 3: Table 1).

The OTUs found in the sampled waterways were distributed amongst 19 phyla (Figure 2). Phyla, absent from Figure 2, were represented by fewer than four genera and/or <50 sequence reads (Glaucophyta, Haptophyta, Foraminifera, Hyphochytriomycota, Bigyra, Amoebozoa and Opisthokonta). The most abundant supergroup with the largest number of OTUs and sequence reads were members of SAR (Stramenopiles, Alveolates and Rhizaria). The phylum with the largest number of identifiable genera was Chlorophyta (89, 7,098 reads), followed by Cil-

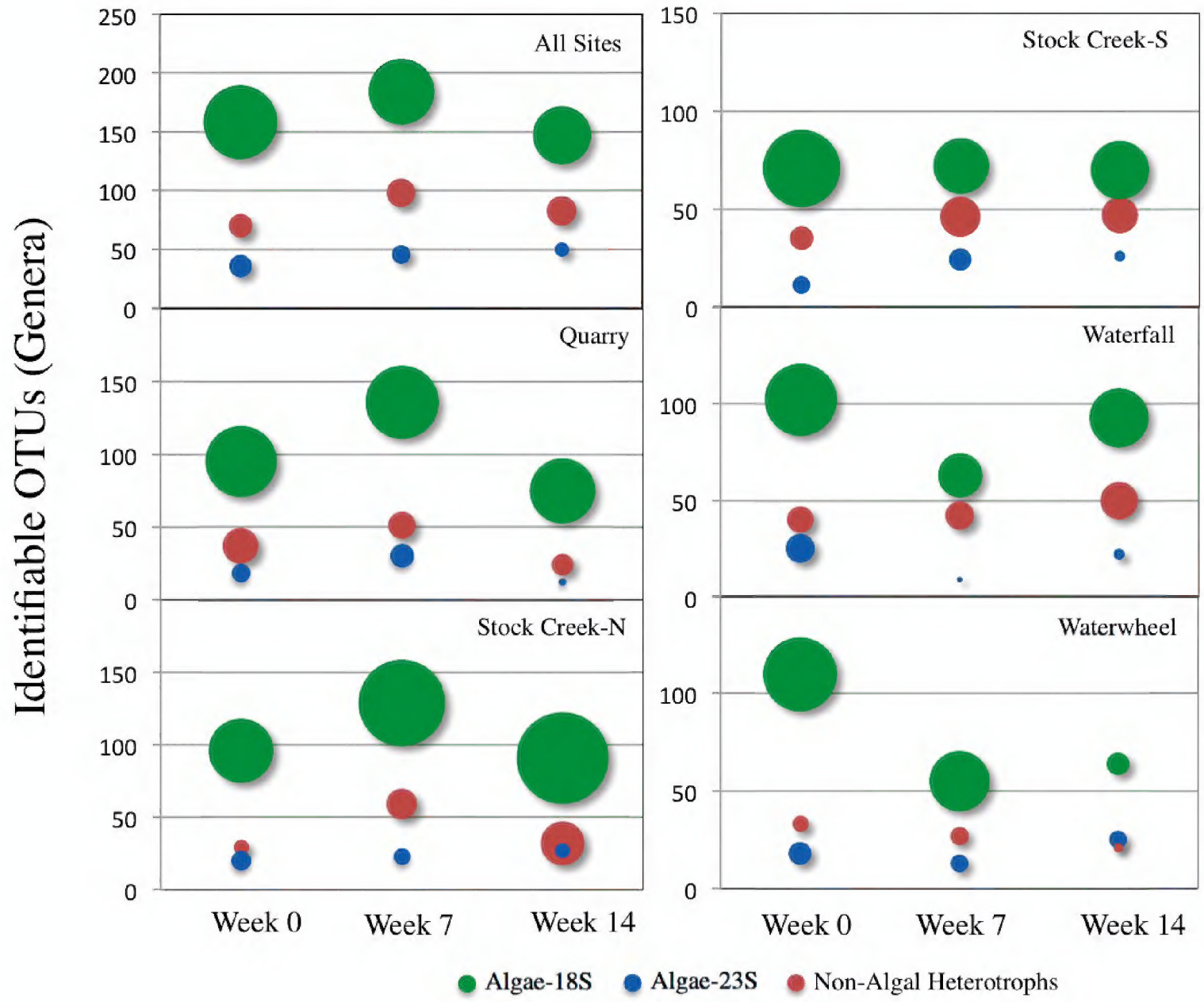

Figure 1. The 18S V4 region barcode identified many more algal OTUs than 23S but the two barcodes lacked concordance. The number of identifiable OTUs (identifiable to the genus level) and sequence abundance of the 18S and 23S barcode sequences were compared. Algal estimates include genera within the phyla Chlorophyta, Glaucophyta, Rhodophyta, Charophyta, Cryptophyta, Haptophyta, Ochrophyta, Bacillariophyta and Dinoflagellata. Heterotroph estimates include genera within the phyla Apicomplexa, Intramacronucleata, Postciliodesmatophora, Cercozoa and Oomycota. The y-axis is the number of genera represented by each circle and the circle diameter represents the number of sequence reads detected for those genera. Green circles represent algal genera detected using 18S, blue represent algal genera detected using 23S and maroon represent heterotrophic genera detected using 18S. The 'all sites' graph represents all genera and sequence reads from the entire study.

iophora (76, 5638), while the most abundant were Bacillariophyta (53 genera, 10,094 reads) and Cryptophyta (8, 9201). Cryptophyta had a large number of identifiable reads (9201) with a relatively small number of genera (8); most of those reads (7449) occurred within the genus *Cryptomonas*. A list of all putative taxa found by this study and the relative abundance of each is listed in Suppl. material 3: Table 1.

Alpha-diversity was estimated and rarefaction curves were produced to determine the success of the sampling strategy (Suppl. material 2: Figure 2). The rarefaction curves suggested that saturation had been approached but not met and that further sampling could be beneficial. They also suggested the Stock Creek samples had the highest taxa richness, the streams flowing into the creek had the next highest richness and the quarry site had the lowest. A C.hat score was calculated for each site and it suggested the taxa richness was near saturation with scores over 0.99 for all sites at the genus and family level (Suppl. material 2: Figure 2). Simpson's (*1-D*) and Shannon-Weiner (*H'*) indices were calculated to estimate diversity for each of the sample sites at each time (Suppl. material 5: Table 3). The Simpson's estimates were very high, ranging from 0.83 to 0.95 for all but the Quarry week 7 sample with 0.41. The Shannon-Weiner index agreed with Simpson's, with high diversity estimates (2.58–3.58) for all but Quarry week 7 (1.3). Taxa evenness was also calculated using Simpson's (E_{1-D}), Shannon-Weiner (H'_{eip}) and *Evar* based indices. The three evenness estimates were low with Simpson's E_{1-D} at <=0.2, *Evar* at <= 0.36 and lower and H'_{eip} ranging from 0.29 to 0.74.

The relative abundance of each genus was calculated to arbitrarily rank read counts as "abundant" (>1% of total reads), "common" (0.95<>0.45%), "uncommon" (0.94<>0.1%) or "rare" (<0.1%) (Figure 3 and Suppl. material 6: Table 4). Twenty-four genera individually comprised ≥1% of the sequences and collectively represented >76% of the total reads and were defined as "abundant". These were in the phyla Bacillarophyta (*Achnanthidium, Gomphonema, Navicula, Nitzschia* and

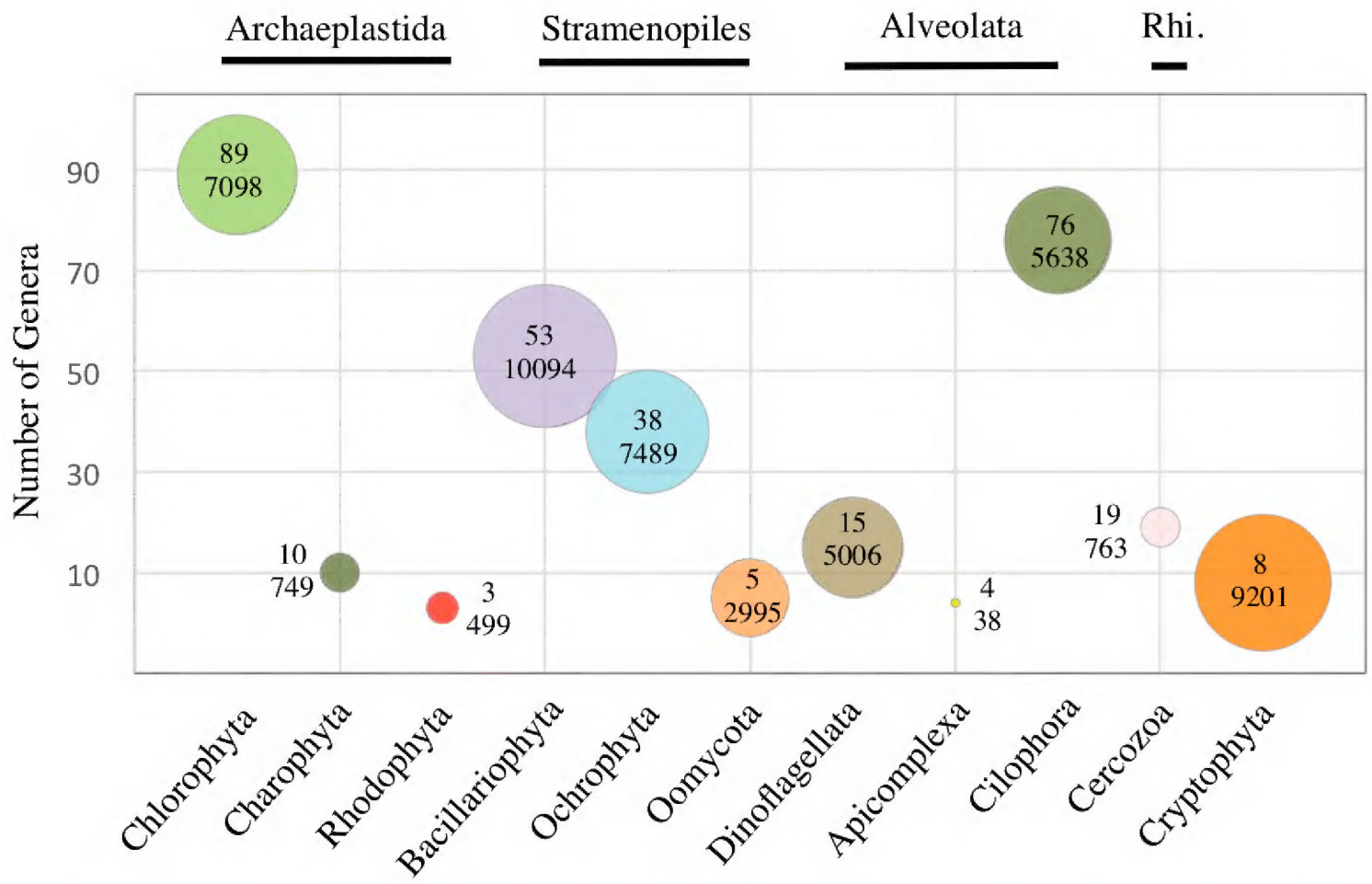

Figure 2. The major protistan phyla detected at the Natural Tunnel State park during the spring-summer of 2016. Each phylum is represented by a circle. The x-axis represents the number of genera identified within each phylum and raw sequence abundance is represented by circle diameter. The number of genera (top number) and raw sequence abundance (bottom number) accompanies each circle. Phyla are organised into the Archaeplastida, Stramenopiles, Alveolata and Rhizaria (Rhi.) super groups according to Adl et al. (2012).

Synedra), Ciliophora (*Tetrahymena, Stylonychia, Helicoprorodon, Spathidium*, *Halteria* and *Stentor*), Miozoa (*Glenodiniopsis, Peridinium, Durinskia* and *Woloszynskia*), Cryptophyta (*Cryptomonas, Rhodomonas* and *Dinobryon*), Chlorophyta (*Desmodesmus* and *Scenedesmus*), two Chrysophyceae (*Spumella* and *Hibberdia*), Phaeophyta (*Ectocarpus*) and Oomycota (*Pythium*).

Spatial location and distribution of taxa (beta-diversity).

The distribution of phyla amongst the five sites is represented in Figure 4A. Chlorophyta, which had the greatest diversity of genera (89), appears to prefer the shallow stagnant water of the lentic quarry over the lotic systems. This preference, however, is skewed due to two highly abundant and common North American genera, *Scenedesmus* and *Desmodesmus*, which had bloomed in the quarry prior to the 22 April sampling date and accounted for 52.7% of the study-wide green algal reads. Bacillariophyta was the most abundant protist phylum with genera primarily located in lotic waterways with low irradiance exposure due to canopy cover. Four were most abundant in the shallow lotic streams (*Achnanthidium*, *Gomphonema*, *Synedra* and *Nitszchia*), while the fifth (*Navicula*) was found at all sample sites with roughly equal abundance. The most abundant genus, *Cryptomonas* (P:Cryptophyta), was relatively abundant in all samples, except the Waterfall and comprised over 15% of the total sequence reads. Seventeen dinoflagellate genera (P: Miozoa) were identified. The phylum Ciliophora was the most evenly distributed with roughly equal numbers of taxa at each site.

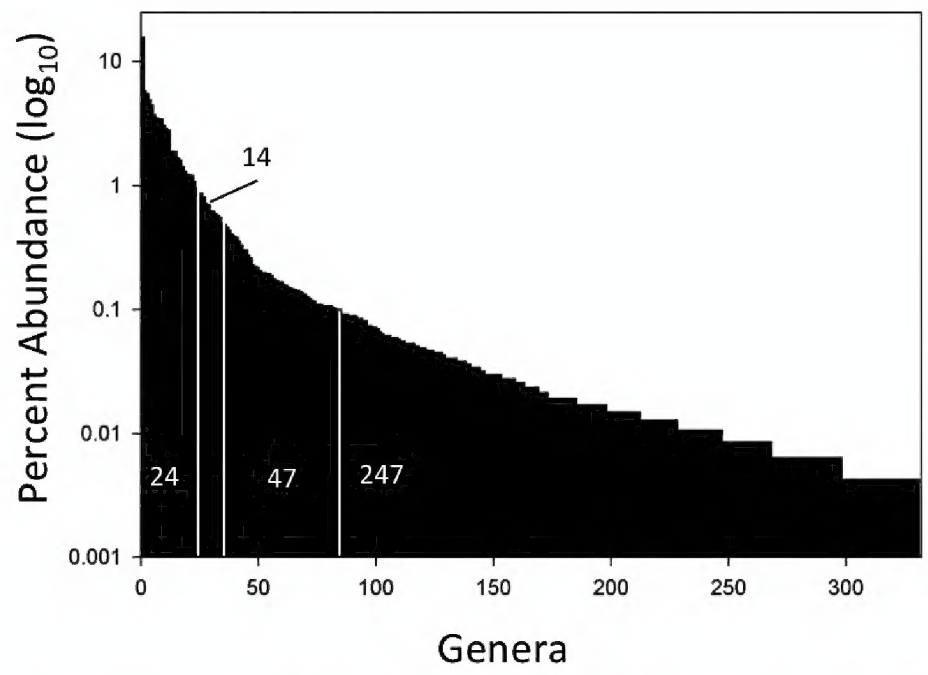

Figure 3. The majority of genera identified in this study were part of the rare microbiome. Rank abundance distribution of genera identified from all sites and times. All genera (x-axis) are arranged according to their relative abundance (percentage) compared to all the identifiable reads (Percent Abundance, x-axis). The vertical white lines divide the genera into four arbitrary abundance categories: "abundant" (>1%), "common" (0.95<>0.45%), "uncommon" (0.94<>0.1%) or "rare" (<0.1%). The numbers associated with each category are the number of genera found in each.

The co-incidence of all genera regardless of abundance was compared between sites (beta-diversity) using Venn diagrams. A comparison of all protist genera shows that 110 (33%) were found at all five sample sites (Figure 4B) and 34 (10.2%) were unique to a single site. When genera were separated into photoautotrophic and primarily heterotrophic (Apicomplexa, Intramacronucleata, Postcil-

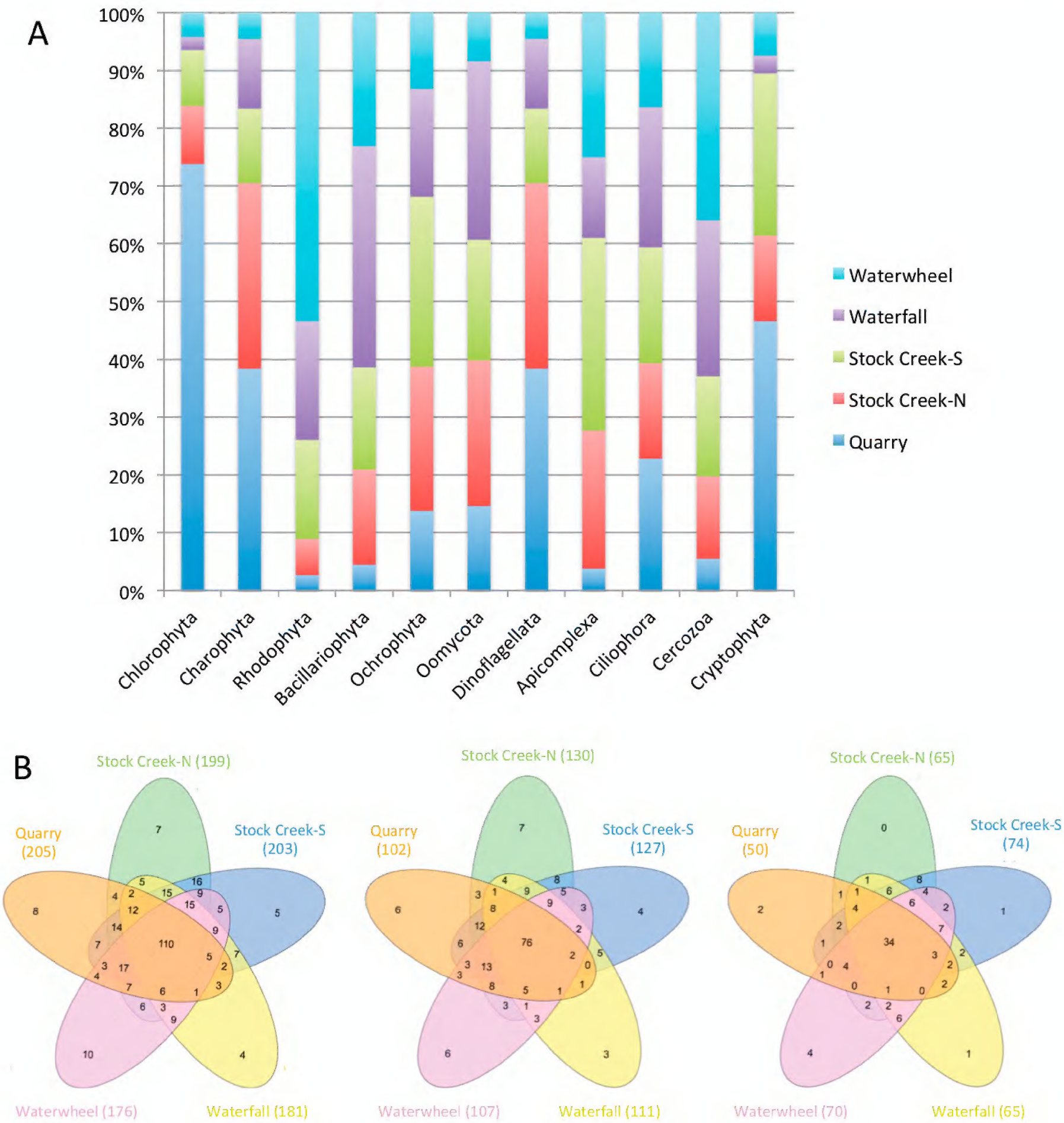

Figure 4. The majority of identifiable genera were found at multiple locations in the park but their relative abundance varies. (**A**) The relative distribution of the phyla at each sample site. (**B**) Three Venn diagrams were prepared to show the distribution of genera amongst the sample sites, represented into groupings of 'all', 'photosynthetic' and 'heterotrophic'. See Figure 1 legend for lists of algal and heterotrophic protist phyla. Diagrams were produced using InteractiVenn (Heberle et al. 2015).

iodesmatophora, Cercozoa and Oomycota) phyla, these proportions changed, with 76 (34%) photoautotrophs and 34 (30.4%) heterotrophs occurring at all sites and 26 (11.8%) photoautotrophs and 8 (7.2%) heterotrophs appearing uniquely at any one site.

Principal coordinates' analyses of genera from the five collections sites and the three ecosystems were also completed. Three distinct profiles were obvious, the Stock Creek sample sites formed a group, the waterfall and waterwheel sites formed a second group and the quarry did not associate with either (Fig. 5A). Next, the photoautotrophic and heterotrophic taxa were separated to see if one or the other had a greater effect on the groupings. The PCO plot with the photoautotrophic taxa only (Fig. 5B) had a very similar distribution as the plot with all protists. The PCO plot, with heterotrophs, varied from all protist plots with the two Stock Creek samples failing to co-locate (Fig. 5C).

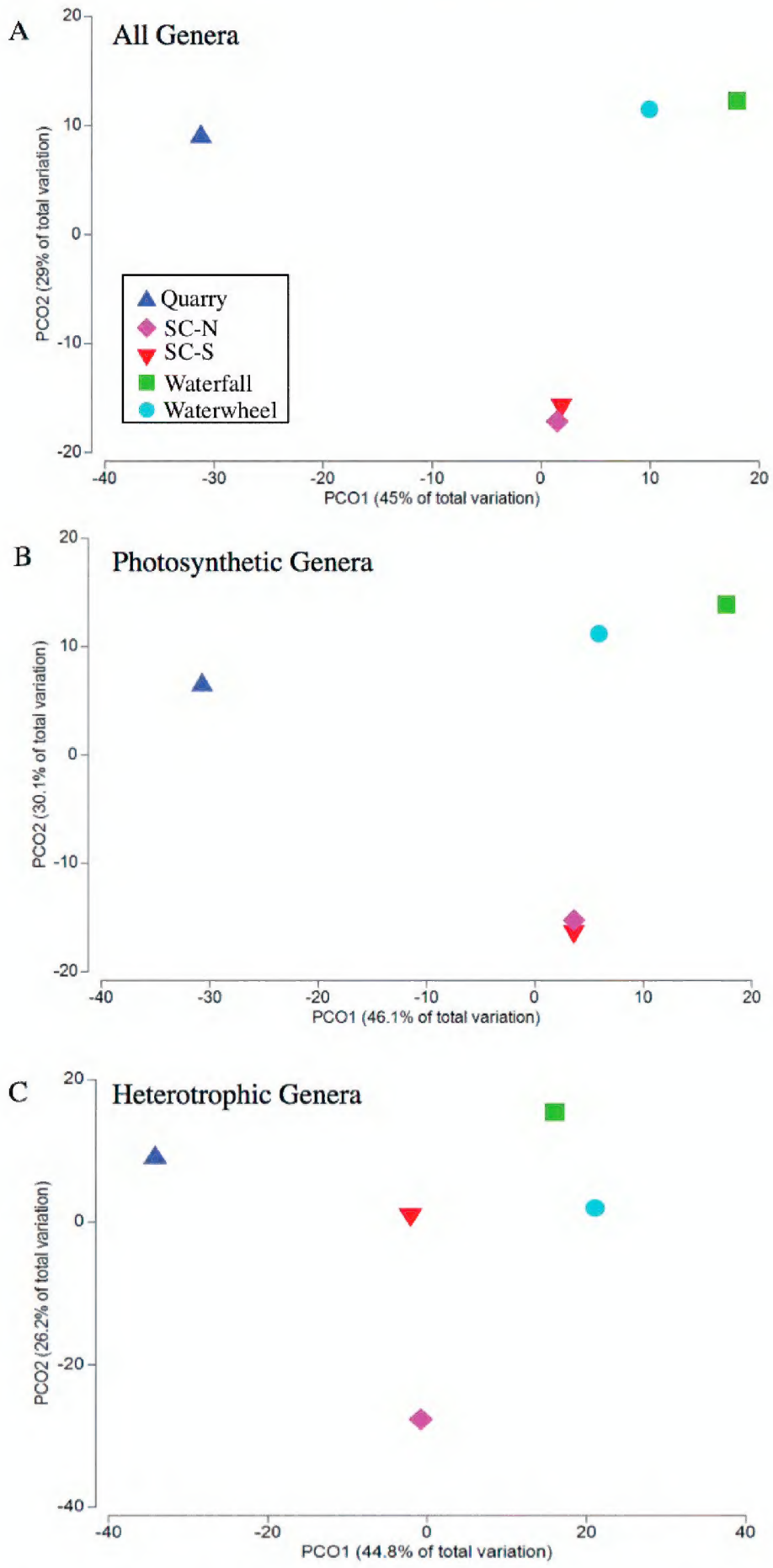

Figure 5. Each sample site has a recognisable protist profile primarily defined by the photosynthetic genera. Principle Co-ordinates Analyses (PCO) was completed to compare the normalised eukaryotic microbiomes of the five sample sites. **A** – all protist genera collected from the five sample sites over the course of the sampling period. **B** – Photosynthetic protist genera. **C** – Heterotrophic protist genera. See Fig. 1 legend for lists of photosynthetic and heterotrophic protist phyla. SC-S and SC-N denote the Stock Creek – South and Stock Creek – North samples, respectively.

Discussion

The use of environmental DNA metabarcoding to identify diversity within eukaryotic microbiomes is a rapidly maturing field and the necessary technologies are becoming more affordable and the bioinformatics pipelines more accessible. In this study, we report its use during a BioBlitz event to describe the diversity of planktonic protists in a state preserve located in the Ridge and Valley physiographic province of the Appalachian region of the United States, an area with reported high biodiversity but with few published surveys of microscopic eukaryotes. Over the course of this study we tested two hypotheses.

Hypothesis 1: The overall protist diversity would be high; perhaps higher than in comparable temperate regions.

We found 3663 OTUs and over 95% of these sequences were identifiable and could be placed in 220 photoautotrophic and 112 heterotrophic genera. It is unknown how many potential taxa are represented by that percentage of unidentifiable sequences but if we consider that 332 identifiable taxa are represented within 84,330 sequences, then that produces an average of 254 sequences per taxa and allows a rough estimate of 15 unidentiable taxa in the 3663 sequences with no match. The small proportion of sequences with no GenBank match is encouraging as it suggests the 18S microbial eukaryotic barcode database may be reaching saturation.

The most abundant taxa found in this study were common North American temperate freshwater protists, such as the green algae *Scenedesmus* and *Desmodesmus* (Shubert and Gärtner 2015), the diatoms *Achnanthidium*, *Gomphonema*, *Synedra*, *Nitszchia* and *Navicula* (Kociolek et al. 2015a and 2015b). The heterotrophic *Cryptomonas* (P:Cryptophyta) was common at all sites and nearly all times where it is presumably serving as a food source for zooplankton (Clay 2015). The phylum Ciliophora was the most evenly distributed with roughly equal numbers of taxa at each site. The most abundant ciliates found in this study, *Halteria*, *Helicoprorodon*, *Spathidium*, *Sylonychia*, *Stentor* and *Tetrahymena* are common freshwater protozoans with diverse diets (Fenchel 1968; Curds et al. 1983; Simon et al. 2008; Zufall et al. 2013).

There were also examples of unexpected taxa. Seventeen dinoflagellate genera (P: Miozoa) were detected in the lotic systems, which is a high number for a freshwater survey (Zalack et al. 2006). Four of these (*Glenodiniopsis, Peridinium, Durinskia and Woloszynskia*), are relatively common North American freshwater genera (Carty and Parrow 2015). Of the remaining 13, six may be new North American records. The most abundant of these, *Kryptoperidinium* is a dinotom (related to *Durinskia*) previously found only in salt and brackish waters (Guiry and Guiry 2017). Another unexpected taxa was *Ectocarpus* (P: Phaeophyta) which appeared in all five sites but was most abundant in the lotic systems. Brown algae are primarily known to occupy marine systems but there are examples of uncharismatic freshwater species from North America (Wehr 2015). An unidentifiable species of brown algae appeared in a metabarcoding survey of the James River (Brown et al. 2015) and has appeared in samples collected by the authors in other parts of the Appalachian region (unpublished data).

The overall distribution of phyla was very consistent with a metabarcoding survey of an inland temperate

freshwater system in France (Simon et al. 2015a). These authors found Crytophytes, Alveolates and Stramenopiles to be highly abundant, including a large representation of dinoflagellate genera. They did not detect taxa within the Archaeplastida at an abundance as high as in our study. A similar, although much longer (2 year) survey of a woodland stream in the Alleghany Plateau region of Ohio, using microscopic identification completed by Zalack et al. (2006), identified 156 algal genera, 52 of which were also found in our survey, suggesting comparably high diversity of algae between this site and ours. Punctate sampling of four freshwater ponds and one brook in France reported 812 OTUs (Simon et al. 2015a), a two year survey of the same freshwater systems, sampled monthly, reported 3,742 (Simon et al. 2015b) and another, focusing on two ponds in the same area that underwent extreme drought, reported 3,132 (Simon et al. 2016). A temporal study, where samples were taken every 2–3 days over the course of a single summer from Lake Geneva, found 991 OTUs (Mangot et al. 2012). Our observation of 3663 OTUs is only exceeded by the two year survey of the ponds in France, which suggests our overall protist diversity estimate is high, regarding temperate freshwater, but is not higher than other comparable bodies of water.

This 'high-protist biodiversity' conclusion is confirmed by the Simpson and Shannon-Wiener biodiversity indices calculated for our survey (Suppl. material 5: Table 3). The Shannon-Wiener index has been calculated for planktonic protists in several freshwater systems based on microscopic surveys and the values generally fall between 0.5 to 3.5 (Beaver and Crisman 1989, Romo and Miracle 1995, Graham et al. 2004, Wang et al. 2004). Our estimates were comparable to the mid to high end of that range, suggesting a comparably high diversity.

We conclude that the diversity of this physiographic province can be considered high and may have a larger number of green algae than other comparable regions but the overall diversity is not unusually high compared to other temperate bodies of fresh water. This conclusion, however, must be tempered by several caveats. The first is the rare biosphere, which has become one of the hallmarks of metabarcoding studies and is defined as OTUs appearing at a very low relative abundance that tend to escape traditional morphological identification (Grattepanche et al. 2014). We chose to set a relative abundance threshold of 0.1% (just over 40 sequence reads) to define rare biosphere components in this survey. At that level, 247 genera (74.4%) were rare, yet collectively, comprised only 5.8% of the overall sequence reads. In other freshwater studies, Mangot et al. (2012) estimated that 99.8% of their taxa could be considered rare and Debroas et al. (2015) estimated 77.2% of the OTUs from two lakes were rare. The sources of these rare DNAs have been the subject of debate and it has been hypothesised that, along with the rare biosphere, these surveys may be detecting molecular debris, dead cells and/or pseudogenes (Grattepanche et al. 2014). There is also no set definition of a rare biosphere, which could explain the large discrepancy between studies. Debroas et al. (2015) tested this by comparing the 18S V4 rDNA barcode region used by most metabarcoding studies to the same rRNA and found that 76% of the rare biospheres were metabolically active and that these comprised 27.5% of the total microbiome they analysed from two lacustrine systems. Simon et al. (2016) found the presence of a diverse microbiome in the sediment of dry ephemeral pools and attributed this to a "spore-bank" that allows rapid proliferation following droughts. Medinger et al. (2010) has suggested these "resting stages" are not identifiable in morphological studies but would be picked up in DNA-based studies. No matter the source, these cells contribute to the biodiversity estimate while comprising a very small portion of the actual biome which could skew the results.

A second caveat is that the overall number of OTUs found during this census is very likely a gross underestimate. A challenge with environmental metabarcoding studies of this type is the choice of primer pairs and the barcode region, which can limit or bias the diversity of organisms observed. We chose the 18S V4 region amplified by 'universal' primers along with 23S primers to identify photoautotrophic organisms and a third diatom specific 18S region. In our study, we found the 'universal' primers provided the greatest number of identifiable barcode sequences and, unexpectedly, identified a nearly identical list of diatom taxa as the diatom specific primers. Previous studies have demonstrated that the universal 18S V4 primer pair can severely underestimate the total number of taxa actually present. Stoeck et al. (2006) found the number of identifiable protist OTUs were nearly doubled when taxa specific primer pairs were used rather than universal primers. Similarly, a survey of soil protists found that 71% of the taxa revealed by specific primers were not identified by the use of universal ones (Lentendu et al. 2014). Geisen et al. (2015) demonstrated that the universal 18S V4 primer pair had a PCR bias for ciliates resulting in over-representation while flagellates and amoebae were underrepresented or missing. This could explain the abundance of ciliates we found and the paucity of amoebae. Although the universal primer pair is imperfect, we used it as an initial survey tool with the hope of parlaying our findings into future studies which focus on a specific group of protists.

A third caveat was our use of an alternative bioinformatics pipeline that used commercially available software and required building a custom searchable database from GenBank rather than using the curated databases PR2 (Guillou et al. 2013) or SILVA. Although our methodology is not the most commonly accepted one, we found the use of a commercial package with a graphical user interface to be attractive and think this could open up metabarcoding to a wider number of users. With this is the understanding that experts have not taxonomically curated the GenBank records as carefully as PR2 and SILVA and there is a possibility of unintended misrepresentation.

Hypothesis 2: The five sample sites have distinct protist profiles.

Five sites were sampled, representing three common waterways found in this region (i) a perennial lotic creek/river, (ii) shallow lotic mountainside streams and (iii) a lentic ephemeral pool. Comparisons of genera from each site found that few taxa were unique to a single location (Figure 4B), yet PCO analysis suggested the three systems had distinct protist biomes, despite their close proximity (Figure 5). A previous study of six freshwater lakes, sampled monthly from April – August using T-RFLP and NGS strategies, suggested that the strongest factor contributing to the differences between the lakes was their geographical distance from one another and not biological or chemical factors (Lepère et al. 2013). Another study of geographically close small freshwater systems found that four close sample sites (2–9.5 km apart) retained distinctive protist profiles over the course of a two year survey, suggesting that environmental or biological factors, not distance, are the primary contributing factors to the protist biome (Simon et al. 2015b). The study area and data presented in our study closely aligns with the latter study suggesting that the protist profiles of these areas are unique so we were unable to reject our hypothesis.

A second observation to stem from the PCO analyses was that the eukaryotic microbiome profiles are likely to be dominated by the autotrophic organisms. The photosynthetic protists were more abundant than heterotrophic ones in the Natural Tunnel environments with a ratio of approximately 6:1. This contradicts a previous molecular survey of freshwater ponds in France where the most abundant protists were planktonic ciliates (Šlapeta et al. 2005) but is concordant with a recent meta-analysis of all publicly available barcode data from freshwater systems where a much larger number of photosynthetic OTUs than heterotrophic ones were identified (Debroas et al. 2017). We hypothesise that the differences in the number of ciliates versus algal protists is likely dependent upon the quickly changing micro-climactic changes or low buffering capacity of the waterways which promote the growth of either algae or bacteria. A study of a continuously flowing looped freshwater feature on the campus of Tsinghua University which could mimic a small stream or creek, demonstrated that the protist communities within the system were significantly different between localised areas and there was a correlation to these changes and the concentration of polyatomic ions (Wang et al. 2004). We are currently performing a long term sampling project of the Natural Tunnel park to test this hypothesis.

Conclusions

The technique of metabarcoding, although imperfect, provides an attractive entry point for the study of microbial eukaryotes biodiversity in small understudied freshwater ecosystems. The 213 families and 332 genera found in this study provide a base-line repository of protist information for a natural area in the Ridge and Valley region of the Appalachian region of North America where protist biodiversity has not received much attention, until now. Over the course of this study, we concluded that the region has a high number of green algae compared to other temperate regions, but not an unusually high overall protist biodiversity. We were also unable to disprove the hypothesis that the three bodies of water would have distinct protist biomes.

Acknowledgements

The authors wish to thank the Natural Tunnel State Park in Scott County, VA for generous access to the park and sampling privileges for the duration of this project. They also thank Bob VanGundy of UVa-Wise, Daniel Wolf of Ohio University and Karen Fawley and Marvin Fawley of the University of Arkansas at Monticello for critically reading the manuscript and providing invaluable insight and inspiration.

References

Aguilar M, Richardson E, Tan B-F, Walker G, Dunfiled PF, Bass DD, Nesbø C, Foght J, Dacks JB (2016) Next-generation sequencing assessment of eukaryotic diversity in oil sands tailings ponds sediments and surface water. J Eukaryot Microbiol 63: 732–743. https://doi.org/10.1111/jeu.12320

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, Heiss A, Hoppenrath M, Lara E, Gall L, Lynn DH, McManus H, Mitchell EAD, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Schoch CL, Smirnov A, Spiegel FW (2012) The revised classification of eukaryotes. J Eukaryot Microbiol 59: 429–493. https://doi.org/10.1111/j.1550–7408.2012.00644.x

Bachy C, Dolan JR, López-García P, Deschamps P, Moreira D (2013) Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. ISME J 2013: 244–255. http://dx.doi.org/10.1038/ismej.2012.106

Beaver J, Crisman TL (1989) Analysis of the community structure of planktonic ciliated protozoa relative to trophic state in Florida lakes. Hydrobiologia 174: 177–184. https://link.springer.com/article/10.1007/BF00008155

Bricheux G, Morin L, Le Moal G, Coffe G, Balestrino D, Charbonnel N, Bohatier J (2013) Forestier, C. Pyrosequencing assessment of prokaryotic and eukaryotic diversity in biofilm communities from a French river. Microbiology Open 2: 402–414. https://doi.org/10.1002/mbo3.80

Brown BL, LePrell RV, Franklin RB, Rivera MC, Cabral FM, Eaves HL, Gardiakos V, Keegan KP, King TL (2015) Metagenomic analysis of planktonic microbial consortia from a non-tidal urban-impacted segment of James River. Stan Genomic Sci 10: 65. https://doi.org/10.1186/s40793-015-0062-5

Cannon MV, Hester J, Shalkhause A, Chan ER, Logue K, Small ST, Serre D (2016) In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversi-

ty surrounding the Cuyahoga River. Sci Rep 6: 22908. https://doi.org/10.1038/srep22908

Caron DA, Countway PD, Savia P, Gast RJ, Schnetzer A, Moorthi SD, Dennett M.R, Moran DM, Jones AC (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Env Microbiol 75: 5797–5808. https://doi.org/10.1128/AEM.00298-09

Carty S, Parrow MW (2015) Dinoflagellates. In: Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Charvet S, Vincent WF, Comeau A, Lovejoy C (2012) Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. Front Microbiol 3: 422. https://doi.org/10.3389/fmicb.2012.00422

Charvet S, Vincent WF, Lovejoy C (2014) Effects of light and prey availability on Arctic freshwater protist communities examined by high-throughput DNA and RNA sequencing. FEMS Microbiol Ecol 88: 550–564. https://doi.org/10.1111/1574-6941.12324

Clarke KR, Gorley RN (2015) PRIMER v7: User Manual/Tutorial. PRIMER-E, Plymouth.

Clay BL (2015) Cryptomonads. In Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Curds CR, Gates MA, Roberts DMcL (1983) British and other freshwater ciliated protozoa Part II Ciliophora: Oligohymenophora and Polyhmenophora. Cambridge University Press.

Debroas D, Hugoni M, Domaizon I (2015) Evidence for an active rare biosphere within freshwater protists community. Mol Ecol 24: 1236–1247. https://doi.org/10.1111/mec.13116

Debroas D, Domaizon I, Humbert J-F, Jardillier L, Lepère C, Oudart A, Taïb N (2017) Overview of freshwater microbial eukaryote diversity: a first analysis of publicly available metabarcoding data. FEMS Microbiol Ecol 93: fix023. https://doi.org/10.1093/femsec/fix023

Downing JA, Prairie YT, Cole JJ, Duarte CM, Tranvik LJ, Striegl RG, Mcdowell WH, Kortelainen P, Caraco NF, Melack JM, Middelburg J (2006) The global abundance and size distribution of lakes, ponds, and impoundments. Limnol Oceanog 51: 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Downing JA (2010) Emerging global role of small lakes and ponds: little things mean a lot. Limnetica 29: 9–24. http://www.limnetica.com/documentos/limnetica/limnetica-29-1-p-9.pdf

Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034–1038. https://doi.org/10.1126/science.1153213

Fenchel T (1968) The ecology of marine microbenthos II. The food of marine benthic ciliates. Ophelia 5: 73–121. https://doi.org/10.1080/00785326.1968.10409626

Finlay BJ, Maberly SC, Cooper JI (1997) Microbial diversity and ecosystem function. OIKOS 80: 209–213. www.jstor.org/stable/3546587

Finlay BJ, Esteban GF (1998) Freshwater protozoa: biodiversity and ecological function. Biodivers Conser. 7: 1163–1186. https://link.springer.com/article/10.1023/A:1008879616066

Geisen S, Laros I, Vizcaíno A, Bonkowski M, De Groot A (2015) Not all are free-living: high-throughput DNA metabarcoding revelas a diverse community of protists parasitizing soil metazoan. Mol Ecol 24: 4556–4569. https://doi.org/10.1111/mec.13238

Graham JM, Kent AD, Lauster GH, Yannarell AC, Graham LE, Triplett EW (2004) Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern temperate humic lake: diversity in a dinoflagellate dominated system. Microb Ecol 48: 528–540. https://link.springer.com/article/10.1007%2Fs00248-004-0223-3

Grattepanche J-D, Santoferrara LF, McManus GB, Katz LA (2014) Diversity of diversity: conceptual and methodological differences in biodiversity estimates of eukaryotic microbes as compared to bacteria. Trends Microbiol 22: 432–437. https://doi.org/10.1016/j.tim.2014.04.006

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvarson B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmerman P, Christen R (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41: D597–604. https://doi.org/10.1093/nar/gks1160

Guiry MD, Guiry GM (2017) *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org; [searched July–August 2017]

Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics 16: 169. https://doi.org/10.1186/s12859-015-0611-3

Hsieh TC, Ma KH, Chao A (2016) iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers) Methods Ecol and Evol 7: 1451–1456. https://doi.org/10.1111/2041-210X.12613

Jack JD, Gilbert JJ (1997) Effects of metazoan predators on ciliates in freshwater plankton communities. J Euk Microbiol. 44: 194–199. https://doi.org/10.1111/j.1550-7408.1997.tb05699.x

John DM (2002) The Freshwater Algal Flora of the British Isles: An Identifiation Guide to Freshwater and Terrestrial Algae. In: John DM, Whitton BA, Brook AJ (Eds) British Phycological Society, Natural History Museum (London, England). Cambridge University Press.

Kammerlander B, Breiner H-W, Filker S, Sommaruga R, Sonntag B, Stoeck T (2015) High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalayan mountains. FEMS Microbiol Eco. 91: fiv010. https://doi.org/10.1093/femsec/fiv010

Kociolek JP, Theriot EC, Williams DM, Julius M, Stoermer EF, Kingston JC (2015a) Centric and araphid diatoms. In: Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Kociolek JP, Spaulding SA, Lowe RL (2015b) Bacillariophyceae: the raphid diatoms. In: Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Lefranc M, Thénot A, Lepère C, Debroas D (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. Appl Environ Microbiol 71: 5935–5942. https://doi.org/10.1128/AEM.71.10.5935-5942.2005

Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9: 119–130. https://doi.org/10.1038/nrmicro2504

Lentendu G, Wubet T, Chatzinotas A, Wilhelm C, Buscot F, Schlegel M (2014) Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. Mol Ecol 23: 3341–3355. https://doi.org/10.1111/mec.12819

Lepère C, Domaizon I, Taïb N, Mangot J-F, Bronner G, Boucher D, Debroas D (2013) Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. FEMS Microbiol Ecol 85: 85–94. https://doi.org/10.1111/1574-6941.12100

López-García P, Rodriguez-Valera F, Pedrós-Alió C, Moreira D (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409: 603–607. https://doi.org/10.1038/35054537

Mangot J-F, Domaison I, Taïb N, Marouni N, Duffaud E, Bronner G, Debroas D (2012) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. Environ Micro 15: 1745–1758. https://doi.org/10.1111/1462-2920.12065

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwa B, Schlötterer C, Boenigk J (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Mol Ecol 19: 32–40. https://doi.org/10.1111/j.1365-294X.2009.04478.x

Moon-van der Staay S, De Wachter R, Vaulot D (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409: 607–610. https://doi.org/10.1038/35054541

Moreira D, López-García P (2002) The molecular ecology of microbial eukaryotes unveils a hidden world. Trends Microbiol 10: 31–38. https://doi.org/10.1016/S0966-842X(01)02257-0

Moss B (2017) Ponds and Small Lakes: Microorganisms and Freshwater Ecology Series: Naturalists' Handbooks Vol 32. Pelagic Publishing, Exeter, U.K.

Neves RJ, Bogan AE, Williams JD, Ahlstedt SA, Hartfield PW (1997) Aquatic Fauna in Peril: The Southeastern Perspective. Special publication 1. In: Benz GW, Collins DE (Eds) Southeast Aquatic Research Institute, Lenz Design and Communications, Decatur, Georgia.

Nolte V, Pandey RV, Jost S, Medinger R, Ottenwälder B, Boenigk J, Schlötterer C (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extant of protist diversity. Mol Ecol 19: 2908–2915. https://doi.org/10.1111/j.1365-294X.2010.04669.x

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2017) vegan: Community Ecology Package. R package version 2.4–3. http://CRAN.R-project.org/package=vegan

Parmalee P, Bogan AE (1998) The freshwater mussels of Tennessee. University of Tennessee Press, Knoxville, Tennessee.

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirků M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EAD, Nitsche F, Romeralo M, Saunders GW, Simpson AGB, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C (2012) CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol 10: e1001419. https://doi.org/10.1371/journal.pbio.1001419

Pielou EC (1969) An introduction to mathematical ecology. Wiley, NY. https://doi.org/10.1002/bimj.19710130308

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41: D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team (2017) R: a language and environment for statistical computing. https://www.R-project.org

Romo S, Miracle MR (1995) Diversity of the phytoplankton assemblages of a polymictic hypertrophic lake. Arch Hydrobio. 132: 363–384.

Sherwood AR, Presting GG (2007) Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. J Phycol 43: 605–608. https://doi.org/10.1111/j.1529-8817.2007.00341.x

Shubert E, Gärtner G (2015) Nonmotile coccoid and colonial green algae. In: Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Sieracki ME, Gobler CJ, Cucci TL, Their EC, Gilg IC, Keller MD (2004) Pico- and nanoplankton dynamics during bloom initiation of *Aureococcus* in a Long Island, NY bay. Harmful Algae 3: 459–470. https://doi.org/10.1016/j.hal.2004.06.012

Simon EM, Nanney DL, Doerder FP (2008) The "Tetrahymena pyriformis" complex of cryptic species. Biodiv Conserv 17: 365–380. https://link.springer.com/article/10.1007/s10531-007-9255-6

Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, López-García P (2015a) Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. Environ Microbiol 17: 3610–3627. https://doi.org/10.1111/1462-2920.12591

Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P, Jardillier L (2015b) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. ISME J 9: 1941–1953. https://doi.org/10.1038/ismej.2015.6

Simon M, López-García P, Deschamps P, Restoux G, Bertolino P, Moreira D, Jardillier L (2016) Resilience of Freshwater Communities of Small Microbial Eukaryotes Undergoing Severe Drought Events. Front Microbiol 7: 812. https://doi.org/10.3389/fmicb.2016.00812

Šlapeta J, Moreira D, López-García P (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc Biol Sci 272: 2073–2081. https://doi.org/10.1098/rspb.2005.3195

Smith B, Wilson JB (1996) A consumer's guide to evenness indices. OIKOS 76: 70–82. https://doi.org/10.2307/3545749

Stern RF, Horak A, Andrew RL, Coffroth M-A, Anderson RA, Kupper FC, Jameson I, Hoppenrath M, Veron B, Kasai F, Brand J, James ER, Keelin PJ (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. PLos One 5: e13991. https://doi.org/10.1371/journal.pone.0013991

Stocker JG (1988) Phototrophic picoplankton: An overview from marine and freshwater ecosystems. Limnol Oceanog. 33: 765–775. https://doi.org/10.4319/lo.1988.33.4part2.0765

Stoeck T, Hayward B, Taylor GT, Varela R, Epstein SS (2006) A multiple PCR-primer approach to access the icroeukaryotic diversity in environmental samples. Protist 157: 31–43. https://doi.org/10.1016/j.protis.2005.10.004

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19 (Suppl. 1): 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Stoeck T, Breiner H-W, Filker S, Ostermaler V, Kammerlander B, Sonntag BA (2014) Morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Env Microbiol 16: 430–444. https://doi.org/10.1111/1462-2920.12194

Stoecker DK, Capuzzo J McD (1990) Predation on protozoa: its importance to zooplankton. J Plank Re. 12: 891–908. https://doi.org/10.1093/plankt/12.5.891

Wang R-C, Wen X-h, Yi Q (2004) Succession of aquatic microbial communities as a result of the water quality variations in continuous water. J Env Sci 16: 772–778.

Wehr JD, Sheath RG, Kociolek JP (Eds) (2015) Freshwater Algae of North America, 2nd Ed. Academic Press.

Wehr JD (2015) Brown Algae. In: Wehr JD, Sheath RG, Kociolek JP (Eds) Freshwater Algae of North America, 2nd Ed. Academic Press.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science 347: 6223. https://doi.org/10.1126/science.1257594

Yoon T-H, Kang H-E, Kang C-K, Lee S-H, Ahn D-H, Park H, Kim H-W (2016) Development of a cost-effective metabarcoding strategy for analysis of the marine phytoplankton community. PeerJ 4: e2115. https://doi.org/10.7717/peerj.2115

Zalack JT, Casamatta DA, Verb RG, Vis ML (2006) A two-year survey of the algal community in a woodland stream from southeastern Ohio. Northeast. Nat. (Stueben) 13: 301–318. https://www.jstor.org/stable/i294418

Zhao B, Chen M, Sun Y, Yang J, Chen F (2011) Genetic diversity of picoeukaryotes in eight lakes differing in trophic status. Can J Microbio. 57: 115–126. https://doi.org/10.1139/w10-107

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. Org Divers Evol 11: 173–192. https://link.springer.com//article/10.1007/s13127-011-0050-6

Zufall RA, Dimond KL, Doerder FP (2013) Restricted distribution and limited gene flow in the model ciliate *Tetrahymena thermophile*. Mol Ecol 22: 1081–1091. https://doi.org/10.1111/mec.12066

Supplementary material 1
Figure 1. Sample collection sites
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: Multimedia.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl1

Supplementary material 2
Figure 2. Rarefaction analysis estimates demonstrate that family and genus collections were approaching saturation
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: Multimedia.
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl2

Supplementary material 3
Table 1. All Detected Taxa
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: species data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl3

Supplementary material 4
Table 2. Primers and PCR Conditions.
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: molecular data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl4

Supplementary material 5
Table 3. Diversity Indices.
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: species data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl5

Supplementary material 6
Table 4. The protist genera identified in Natural Tunnel State Park organised by read count.
Authors: A. Bruce Cahoon, Ashley G. Huffman, Megan M. Krager, Roseanna M. Crowell
Data type: species data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.26939.suppl6